Strathprints Institutional Repository

 doi:10.1093/bfgp/els055

Editorial

Complex biomolecular networks: challenges and opportunities

Complex biomolecular networks include a series of networked complex systems ranging from genomic and transcriptomic to proteomic and metabolomic ones. The use of the term 'biological' networks for this type of systems, which is common in the literature, is inappropriate due to the fact that biomolecular networks are a subset of the biological ones, which also include cellular, anatomical or functional networks. After the pioneering works in this area at the end of the 1990s there has been an explosion of results in the application of network techniques to discovering biologically meaningful properties of biomolecular systems.

The simplest paradigm for the study of complex biomolecular systems consists in the fact that these systems are representable as networks in which the nodes represent the entities of the complex system, such as genes, proteins, metabolites, etc., and the links represent a variety of possible interactions among them, which includes physical interactions or chemical transformations. The first challenge is then the problem of how to capture the vast complexity of a biomolecular system by a simplified representation in the form of a network. In most of the cases, the biomolecular systems are represented as simple networks or pseudonetworks. These are networks having only sets of nodes and links (and possible self-loops) but no directionality or weights are used for the links. This kind of representation is useful when we are interested in analysing the influence of the 'pure' topology of the network on the biological processes taking place on the systems under study. The addition of directionality is necessary when studying transformations of some entities into others, such as in the case of metabolic networks. However, even the use of directed networks is not the best choice for representing the whole complexity of a metabolic (or any other reaction) network. As an example we mention the many problems that have arisen when using the identification of shortest paths in metabolic networks as possible metabolic pathways. The problem is that a compound–compound projection of a reaction network is not appropriate for selecting biologically meaningful reaction paths. The use of weights for the links has ameliorated the problem but it is still not optimal for representing such complex systems. This problem has been largely ignored by researchers in this area and has produced strong criticisms to network theory as a tool for studying metabolic networks [1]. However, as we have remarked elsewhere [2] the problem here is not that *network techniques are wrong, but that wrong questions have been asked to a representation of a system which is not appropriate for such question*. The opportunities here arise by use of better representations of the complex biomolecular systems. The first obvious candidate is the use of directed hypernetwork representation for the metabolic system as has been already proposed with very promising results [3]. A similar situation appears also when studying gene regulatory networks as commented by Klein *et al*. in the first review of this issue.

Another challenging area of research for the study of complex biomolecular networks is that of the scaling of their degree distributions. As it is widely documented today, there are many real-world systems for which there are significant deviations from a 'democratic' distribution of node degrees expected from the 'classical' random networks. Instead they display a very 'egoistic' type of distribution, in which very few nodes have very large degree and most of the nodes has relatively low degree. Initially, a lot of 'evidences' were collected about the existence of power-law degree distribution in biomolecular networks. The interpretation and in some way exaggerations about the role of this type of distributions in biological terms was criticized elsewhere [4]. More accurate analysis has revealed that in fact many of these networks do not display power-law degree distribution but other type of the so-called fat-tailed distributions. For instance, Stumpf and Ingram [5] analysed a few protein–protein interaction networks

discovering that they have degree distributions of the types of stretched exponential, lognormal, Weibull and gamma distributions. Now, if the irregularity in the degree distribution of a network means something relevant from a biological point of view the challenge here is how to compare such irregularity from the degree distribution. If we have such a large zoo of distributions, all of them with very different mathematical shapes, it is far from trivial how to determine a ranking of degree heterogeneity. The opportunities in this area arise from the use of algebraic network techniques. For instance, a degree heterogeneity index has been produced on the basis of quadratic forms of the Laplacian matrix of a network that has a nice spectral representation and allows a ranking of any network according to their degree of heterogeneity [6]. Some other spectral methods for visualization and analysis of heterogeneous networks are reviewed by McDonald *et al*. in the current issue. Other structural approaches are described by Gáspár and Csermely on the basis of a detailed account on combinatorial rigidity analysis and its relation to protein structures and properties. The field of networks representing protein structures is reviewed in two independent reviews in this issue by Greene and by Atilgan and Atilgan, respectively, where it is concluded that viewing proteins as network systems give significant insights into the determinants of protein structure, stability and folding. The reviews of de las Rivas and Fontanillo as well as that of Jordan *et al*. account for the network techniques used for the analysis of protein–protein interaction networks. These reviews are followed by the analysis of post-transcriptional regulatory networks, where Chandra Janga shows the role played by a class of proteins called RNA-binding proteins as well as a number of small RNAs in the metabolism of RNAs.

Another challenging area for the study of complex biomolecular networks is that of developing biologically meaningful network dynamical models. Two kinds of dynamics can be easily identified here. The first corresponds to the dynamics of the network as an evolutionary system. The second to the dynamical processes taking place on the nodes and links of these networks. In the first area the challenge is how to incorporate biologically useful information into the dynamical processes in order to reproduce many of the topologically relevant characteristics of these networks. The first types of models developed in this area were designed to reproduce very specific topological characteristics of a network but not a group of properties of relevance in biology. In the second area there has been more success due to the availability of several approaches that include kinetic metabolic models using ordinary differential equations now integrated in the network context or the use of Boolean networks. Some of these techniques are analysed in the reviews of Klein *et al*., Ma and Gao, as well as in the one of Cheng *et al*., where the relation between dynamics and cancer is reviewed.

The final goal of network analysis of complex biomolecular systems is to elucidate some of the mechanisms giving rise to human diseases. This is the topic of the review papers of Janjić and Pržulj and the one by Goh and Choi, where the role played by network topology in understanding biological functions and human diseases is illustrated. The necessity of including dynamical aspects to the study of networks is emphasized in these two reviews as well as in others covered in this issue. That is why the last review of this special issue deals with precisely this problem in the analysis of one of the most complex diseases we confront today: cancer. Cheng *et al*. show how analysing the dynamics of cancer networks can facilitate a better understanding of this disease, helping to design better cancer diagnosis tools, improving the efficiency of biomarker discovery and increasing the accuracy of identifying the oncogenic genes.

New techniques are emerging in the field of network theory that are of fundamental importance for the study of complex biomolecular networks, their dynamics and their role in human diseases. For instance, the emerging area of complex multiplexes, where a system is represented by several layers of networks vertically interconnected is one of the most promising ones. The next challenge is then to integrate all these systems into a theory that allows us to have predictive models of biological functions and diseases in an affective way. As we have seen in this special issue: *new challenges give rise to new opportunities*. Thus, the new challenges in the area of studying complex biomolecular network will give rise to new opportunities for collaborations and integration in this multidisciplinary area of research.

Ernesto Estrada
Department of Mathematics and Statistics
Institute of Complex Systems, University of Strathclyde
26 Richmond Street, Glasgow G11XQ, UK
E-mail: ernesto.estrada@strath.ac.uk

References

1. Bourguignon P-Y, van Helden J, Ouzounis C, *et al*. Computational analysis of metabolic networks. In: Frishman D, Valencia A (eds). *Modern Genome Annotation: the BioSapiens Network*. Springer, 2008.
2. Estrada E. *The Structure of Complex Networks*. Oxford University Press, 2011.
3. Klamt S, Haus U-U, Theis F. Hypergraphs and cellular networks. *PLoS Comput Biol* 2009;**5**:e1000385.
4. Keller EF. Revisiting 'scale-free' networks. *BioEssay* 2005; **27**:1060–8.
5. Stumpf MPH, Ingram PJ. Probability models for degree distributions of protein interaction networks. *Europhys Lett* 2005;**71**:152–8.
6. Estrada E. Quantifying network heterogeneity. *Phys Rev E* 2010;**82**:066102.